



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2016

---

## Recent advances in probabilistic species pool delineations

Karger, Dirk Nikolaus

**Abstract:** A species pool is the set of species that could potentially colonize and establish within a community. It has been a commonly used concept in biogeography since the early days of MacArthur and Wilson's work on Island Biogeography. Despite their simple and appealing definition, an operational application of species pools is bundled with a multitude of problems, which have often resulted in arbitrary decisions and workarounds when defining species pools. Two recently published papers address the operational problems of species pool delineations, and show ways of delineating them in a probabilistic fashion. In both papers, species pools were delineated using a process-based, mechanistical approach, which opens the door for a multitude of new applications in biogeography. Such applications include detecting the hidden signature of biotic interactions, disentangling the geographical structure of community assembly processes, and incorporating a temporal extent into species pools. Although similar in their conclusions, both 'probabilistic approaches' differ in their implementation and definitions. Here I give a brief overview of the differences and similarities of both approaches, and identify the challenges and advantages in their application.

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-125126>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Karger, Dirk Nikolaus (2016). Recent advances in probabilistic species pool delineations. *Frontiers of Biogeography*, 8(2):e30545.

**Title:**

Recent advances in probabilistic species pool delineations

**Journal Issue:**

[Frontiers of Biogeography, 8\(2\)](#)

**Author:**

[Karger, Dirk Nikolaus](#), Department of Systematic and Evolutionary Botany, University of Zurich

**Publication Date:**

2016

**Permalink:**

<http://escholarship.org/uc/item/19c9h4j0>

**Acknowledgements:**

I thank the Swiss National Fund (SNF) for funding (grant: 148691), Michael Kessler for valuable comments on the first draft of the paper, and Guy Atchison for correcting my language.

**Keywords:**

species pools, community composition, community assembly, dark diversity, source pool

**Local Identifier:**

fb\_30545

**Abstract:**

A species pool is the set of species that could potentially colonize and establish within a community. It has been a commonly used concept in biogeography since the early days of MacArthur and Wilson's work on Island Biogeography. Despite their simple and appealing definition, an operational application of species pools is bundled with a multitude of problems, which have often resulted in arbitrary decisions and workarounds when defining species pools. Two recently published papers address the operational problems of species pool delineations, and show ways of delineating them in a probabilistic fashion. In both papers, species pools were delineated using a process-based, mechanistical approach, which opens the door for a multitude of new applications in biogeography. Such applications include detecting the hidden signature of biotic interactions, disentangling the geographical structure of community assembly processes, and incorporating a temporal extent into species pools. Although similar in their conclusions, both 'probabilistic approaches' differ in their implementation and definitions. Here I give a brief overview of the differences and similarities of both approaches, and identify the challenges and advantages in their application.

**Copyright Information:**





**eScholarship**  
University of California

eScholarship provides open access, scholarly publishing services to the University of California and delivers a dynamic research platform to scholars worldwide.

## perspective

## Recent advances in probabilistic species pool delineations

Dirk Nikolaus Karger<sup>1</sup>

<sup>1</sup>Department of Systematic and Evolutionary Botany, University of Zurich, 8008 Zurich, Switzerland; [dirk.karger@systbot.uzh.ch](mailto:dirk.karger@systbot.uzh.ch)

**Abstract.** A species pool is the set of species that could potentially colonize and establish within a community. It has been a commonly used concept in biogeography since the early days of MacArthur and Wilson's work on Island Biogeography. Despite their simple and appealing definition, an operational application of species pools is bundled with a multitude of problems, which have often resulted in arbitrary decisions and workarounds when defining species pools. Two recently published papers address the operational problems of species pool delineations, and show ways of delineating them in a probabilistic fashion. In both papers, species pools were delineated using a process-based, mechanistical approach, which opens the door for a multitude of new applications in biogeography. Such applications include detecting the hidden signature of biotic interactions, disentangling the geographical structure of community assembly processes, and incorporating a temporal extent into species pools. Although similar in their conclusions, both 'probabilistic approaches' differ in their implementation and definitions. Here I give a brief overview of the differences and similarities of both approaches, and identify the challenges and advantages in their application.

**Introduction**

The species pool is a common concept used in theoretical and applied studies in biogeography (Ricklefs 1987, Cornell and Harrison 2014). A species pool is the set of species that could potentially colonize and establish within a community (Zobel 1997, Zobel et al. 1998, Pärtel et al. 2011a, Lessard et al. 2012). Despite such an initially simple, appealing definition, an operational application of a species pool is bundled with a multitude of problems, often resulting in arbitrary decisions, such as fixed geographical extents of regional pools, or thresholds on environmental parameters. These fixed extents and thresholds are often applied simply because we lack a quantitative mechanism to ascertain which species belong to the species pool and which do not. Does a spatial extent of a few km<sup>2</sup> surrounding a site capture all species which can potentially occur in it? Is an environmental threshold of 5°C sufficient to decide whether a species belongs to the species pool or not? Do we simply use all species for which we have data? These questions accumulate when researchers are faced with the need to delineate a species pool, and so does the associated problem

of not having an empirical answer to many of them.

While carefully selected thresholds can be appropriate for some research questions (Lewis et al. 2016), arbitrary choices are generally undesirable, and can even have severe effects on the outcome of studies of community-assembly processes (Kraft et al. 2007, Kissling et al. 2012, Eiserhardt et al. 2013, Karger et al. 2015, Lessard et al. 2016). The definition of a species pool, itself, suggests that in practice species pools should be treated probabilistically. This is an important point, which has been made by several authors before (e.g. Ewald 2002, Karlson & Cornell 2002, Graves & Rahbek 2005), but unfortunately has often been ignored in biogeographical studies because of the practical problems this would create when actually applying it.

Two recent papers (Karger et al. 2016, Lessard et al. 2016), however, pick up the practical challenges of avoiding arbitrary decisions when delineating species pools. They highlight how species pools can be delineated in a probabilistic fashion, how they can be used to detect biotic interactions (Lessard et al. 2016), disentangle the geo-

graphical structure of community-assembly processes (Karger et al. 2016), or investigate time-related processes of community assembly (Karger et al. 2016).

Although similar in their conclusions, both ‘probabilistic approaches’ differ in their implementation and definitions. These distinctions are important if probabilistic species pools are implemented in the future, and here I therefore give a short overview of the differences and similarities of both approaches.

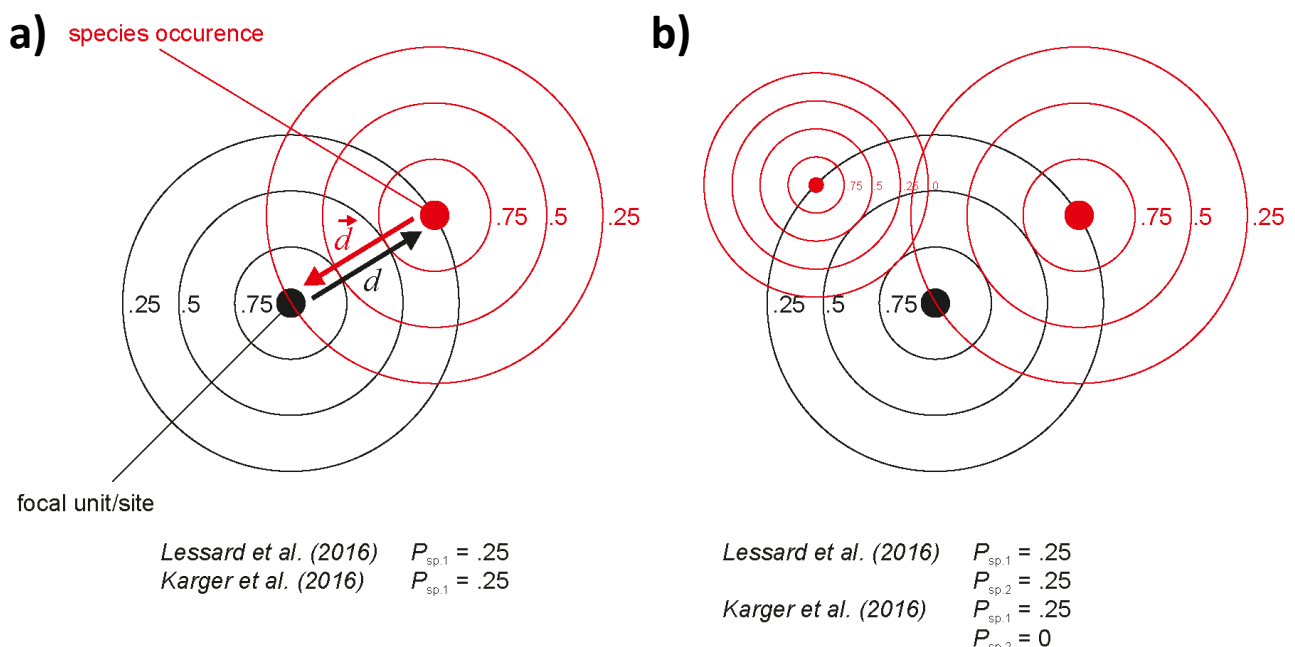
### Site-specific vs. species-specific approaches

Both approaches (Lessard et al. 2016, Karger et al. 2016) are probabilistic, but differ in the way the probabilities are calculated. The method used by Lessard et al. (2016) can be called a ‘site-specific’ approach, while that adopted by Karger et al. (2016) is a ‘species-specific’ one. A ‘site-specific approach’ should not be confused with a ‘site-specific species pool’; species pools are always linked to a focal site, by definition. In the approach used by Lessard et al. (2016), all probabilities are calculated based on a directional vector

originating at the site (Fig 1a). To do so, Lessard et al. (2016) use the distance from a focal site to define probabilities of dispersal. The further away another site is, the lower the probability that a species occurring in this distant site is part of the species pool of the focal site. Therefore, all species with an equal distance to the focal site have the same probability to be part of the species pool.

The approach by Karger et al. (2016) differs as it uses a ‘species-specific approach’. Probabilities of dispersal ( $P_{DD}$ ) are in this case given by a dispersal kernel centred on each occurrence of a species. In this approach, species which have the same distance from a focal site (or focal unit, *sensu* Karger et al. 2016), do not necessarily have the same probability to be part of the species pool. Their probability instead depends on a species-specific dispersal rate, estimated using dispersal traits or simply set to a fixed distance per year.

Both approaches certainly have their advantages and drawbacks. Karger et al.’s (2016) approach ideally needs very detailed information on dispersal rates for each species that could oc-



**Figure 1.** Comparison of a site-specific (black) and species-specific (red) approach. **(a)** A site-specific approach is based on a fixed distance from a focal site to an occurrence ( $d$ ), while **(b)** a species-specific approach is based on a dispersal vector ( $\vec{d}$ ). If  $d$  and  $\vec{d}$  are the same, both approaches will result in the same probability for a species to be part of the species pool. If, however, two species with different dispersal abilities are considered (here shown as different dispersal kernels), the probabilities will differ.

cur in a focal unit, information which is often not available. Even though dispersal rates are known for a few species, and are available via a few databases (e.g. the TRY<sup>1</sup> and D3<sup>2</sup> databases for plant traits), information on actual dispersal rates is typically sketchy. A practical alternative, for the time being, is to estimate dispersal from related morphological or ecological traits (Tamme et al. 2014), to circumvent this problem.

Using a fixed distance from a focal site, as Lessard et al. (2016) did, is therefore an easy workaround of this problem, and certainly can be justified in specific cases. In the case of Lessard et al. (2016), who studied hummingbirds – a rather homogeneous taxonomic group with comparable dispersal traits – equal dispersal probabilities between species might be appropriate. This problematic, however, if such an approach is extended to more than one taxonomic group with heterogeneous dispersal traits. Mathematically, both approaches will result in the same probabilities if fixed distances are used (Fig. 1a). However, they differ as soon as different dispersal rates per species are assumed (Fig. 1b).

### **Integrating phylogenies (Lessard et al. 2016) and time frames (Karger et al. 2016)**

Despite several similarities of both publications, they also contain unique aspects. The integration of probabilistic species pools into an analysis of phylogenetic community composition is certainly the most innovative part of the Lessard et al. (2016) paper. They were able to show that shifting the applied environmental threshold on the species pool can severely influence the inference which can be drawn from phylogenetic community patterns. This important contribution shows that we have to rethink a decade of analysing community phylogenetic patterns.

Karger et al. (2016) did not mention phylogenies, but introduced explicit time frames for which species pools are defined. Although time only entered their equations via the rates of dispersal over time, a full integration into environmental and biotic factors could be achieved in a

similar manner. The quantification of time-related processes such as exact dispersal rates, changes in environmental conditions and biotic interactions might pose the largest challenge in applying their approach in the future. However, it also opens up a wide range of possibilities to investigate the influence of time-related processes on the size and composition of species pools, which have so far been elusive.

### **Detecting biotic processes**

Detecting biotic processes in ecological communities is at best challenging, as it usually requires detailed observational data for a large amount of species, or a long-term experimental setup. A method to detect such biotic processes without directly measuring them might therefore be a neat workaround of the time consuming exercise of directly measuring them (e.g. de Bello et al. 2012).

Lessard et al. (2016) indirectly detected a signal of biotic processes in phylogenetic community patterns. By first accounting for environmental filtering, they constructed a null community in which the effects of environmental filtering are already accounted for. The subsequent analysis revealed patterns resembling those expected under competition. Although this procedure does not get around the common problem of inferring processes from patterns of phylogenetic relatedness, it does illustrate how species pools can be used to disentangle processes from each other in a comparably simple and practicable way.

The Karger et al. (2016) paper argues in a similar direction, although using a simulated approach based on a mechanistic niche model. By simulating differently nested species pools (a dispersal pool, a dispersal + environmentally filtered pool, and a dispersal + environment + biotically filtered pool), they show that increasing the constraint on species pools increasingly lets species pools reassemble realized assemblages. The remaining discrepancies between realized communities and species pools are then attributed to the processes not included in the species pool deline-

<sup>1</sup> [www.try-db.org](http://www.try-db.org), last accessed 30<sup>th</sup> June 2016

<sup>2</sup> [www.seed-dispersal.info/](http://www.seed-dispersal.info/), last accessed 30<sup>th</sup> June 2016



ation (e.g. if the species pool includes dispersal and environment, then the remaining variance can be attributed to biotic processes). This is a line of argument, found within both papers, that is similar to the concept of ‘dark diversity’ (Mokany and Paini 2011, Pärtel et al. 2011a, 2011b, 2013, Ronk et al. 2015) – whatever remains in differences between realized and simulated community structure can be attributed to the processes not included in a species pool delineation.

### Geographical patterns of community assembly processes

Similar to the concepts used to detect biotic processes in a single assemblage, both papers also show the geographical distribution of different processes. From Lessard et al. (2016) a geographical distribution of biotic processes can be inferred from the differences in phylogenetic clustering after accounting for environmental filtering. Increasing the strength of, for example, temperature filtering on the species pool (going from a 0<sup>th</sup> quantile to the 99<sup>th</sup> quantile) shows that biotic processes are not independent of the elevational distribution of the sampled assemblages (Fig. 2a). Karger et al. (2016) showed the geographical distribution of assembly processes by highlighting the difference between different filtered pools and realized community structure for a gridded dataset. They calculated the area under the receiver operating characteristic curve (AUC) to characterize where dispersal and environmentally filtered species pools of Ranunculaceae described realized assemblage structure to a large degree (Fig. 2b, orange). The remaining variance can be then attributed to processes not included in the species pool definition (arguably biotic processes) (Fig. 2b, blue), allowing, for the first time, the study of the geographical distribution of different assembly processes.

### Concluding remarks

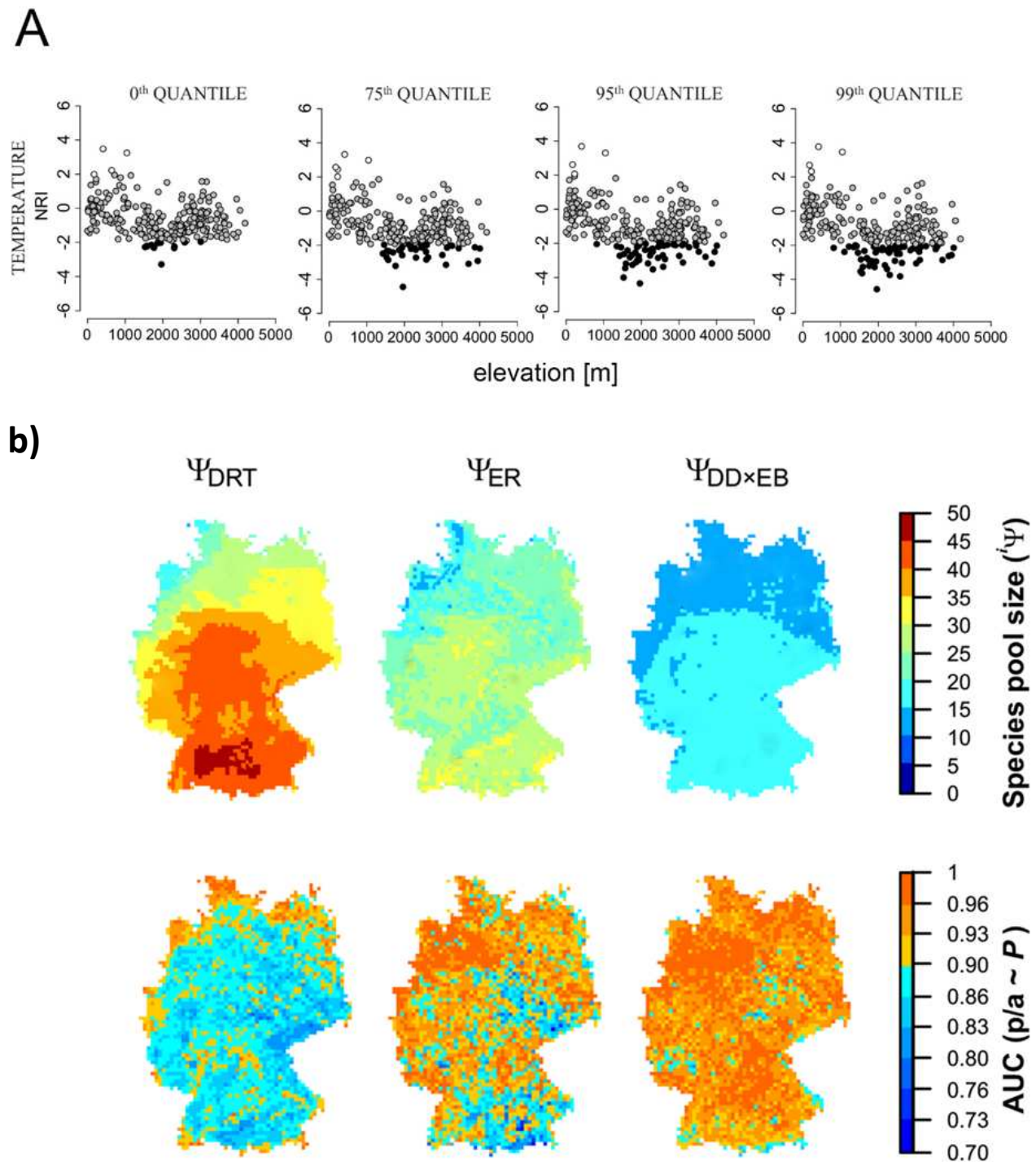
The recurring problem of avoiding arbitrary choices in species pool delineations can, to a large degree, be avoided by using a functional (de Bello et al. 2012) and probabilistic species pool (Ewald 2002, Lessard et al. 2016, Karger et al. 2016).

However, the application of such an approach is mainly hindered by the availability of the data and therefore there are still few methods which can be applied. The uses of probabilistic species pools, such as the detection of biotic processes, their geographical distribution, and the possibility of creating time-dependent species pools, could easily outweigh the efforts needed for delineating probabilistic pools.

An important distinction is that, although both the approaches presented here are ‘probabilistic’, they differ in the way they are calculated; in many cases the difference between ‘site-specific’ and ‘species-specific’ approaches can lead to different probabilities of species pool membership. One needs to decide which approach fits better into a given study system. The two papers focused on herein represent a way out of the myriad arbitrary ways to delineate species pools. Although still difficult to implement, the currently increasing amount of species-specific data calls for increasing application of probabilistic frameworks instead of the still commonly used arbitrary delineations of species pools.

### References

- de Bello, F., Price, J.N., Münkemüller, T. et al. (2012) Functional species pool framework to test for biotic effects on community assembly. *Ecology*, 93, 2263–2273.
- Cornell, H.V. & Harrison, S.P. (2014) What Are Species Pools and When Are They Important? *Annual Review of Ecology, Evolution, and Systematics*, 45, 45–67.
- Eiserhardt, W.L., Svenning, J.-C., Baker, W.J., Couvreur, T.L.P. & Balslev, H. (2013) Dispersal and niche evolution jointly shape the geographic turnover of phylogenetic clades across continents. *Scientific Reports*, 3, 1164.
- Ewald, J. (2002) A probabilistic approach to estimating species pools from large compositional matrices. *Journal of Vegetation Science*, 13, 191–198.
- Graves, G.R. & Rahbek, C. (2005) Source pool geometry and the assembly of continental avifaunas. *Proceedings of the National Academy of Sciences USA*, 102, 7871–7876.
- Karger, D.N., Tuomisto, H., Amoroso, V.B., Darnaedi, D., Hidayat, A., Abrahamczyk, S., Kluge, J., Lehnert, M. & Kessler, M. (2015) The importance of species pool size for community composition. *Ecography*, 38, 1243–1253.
- Karger, D.N., Cord, A.F., Kessler, M. et al. (2016) Delineating probabilistic species pools in ecology and biogeography. *Global Ecology and Biogeography*, 25, 489–501.
- Karlson, R.H. & Cornell, H.V. (2002) Species richness of coral assemblages: Detecting regional influences at local spatial scales. *Ecology*, 83, 452–463.



**Figure 2.** Two examples of how the elevational and geographical distribution of community assembly processes can be inferred from probabilistic species pools. **(a)** Using a variety of thresholds (using quantiles) of the probabilities to include or exclude species from the pool, shows where biotic processes are operational. The investigated hummingbird assemblages in the lowlands below 1000 m elevation do not seem to be affected by biotic processes, while those above 1000 m become phylogenetically more even (lower NRI) when the environmental filtering is increased (black circles). **(b)** Delineating different probabilistic species pools for Ranunculaceae in Germany and comparing them with realized communities can indicate the areas in which the processes not included in the species pool delineations might be important for community assembly (indicated by low AUC values,  $p/a \sim P$  = present/absent ~ Probability). Abbreviations:  $\Psi$  = probabilistic species pool, process subscripts: DRT = dispersal based suitable dispersal pathways, DD = dispersal based on a fixed distance, ER = environmental suitability based on niche modelling, EB = environmental suitability based on Beals smoothing, x denotes combinations of factor groups.



- Kissling, W.D., Dormann, C.F., Groeneveld, J. et al. (2012) Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography*, 39, 2163–2178.
- Kraft, N.J., Cornwell, W.K., Webb, C.O. & Ackerly, D.D. (2007) Trait evolution, community assembly, and the phylogenetic structure of ecological communities. *The American Naturalist*, 170, 271–283.
- Lessard, J.-P., Belmaker, J., Myers, J.A., Chase, J.M. & Rahbek, C. (2012) Inferring local ecological processes amid species pool influences. *Trends in Ecology & Evolution*, 27, 600–607.
- Lessard, J.-P., Weinstein, B.G., Borregaard, M.K., Marske, K.A., Martin, D.R., McGuire, J.A., Parra, J.L., Rahbek, C. & Graham, C.H. (2016) Process-Based Species Pools Reveal the Hidden Signature of Biotic Interactions Amid the Influence of Temperature Filtering. *The American Naturalist*, 187, 75–88.
- Lessard, J.-P., Weinstein, B.G., Borregaard, M.K. et al. (2016) Process-based species pools reveal the hidden signature of biotic interactions amid the influence of temperature filtering. *The American Naturalist*, 75–88.
- Lewis, R.J., Szava-Kovats, R. & Pärtel, M. (2016) Estimating dark diversity and species pools: an empirical assessment of two methods. *Methods in Ecology and Evolution*, 7, 104–113.
- Mokany, K. & Paini, D.R. (2011) Dark diversity: adding the grey. *Trends in Ecology & Evolution*, 26, 264–265.
- Pärtel, M., Szava-Kovats, R. & Zobel, M. (2011a) Dark diversity: shedding light on absent species. *Trends in Ecology & Evolution*, 26, 124–128.
- Pärtel, M., Szava-Kovats, R. & Zobel, M. (2011b) Discerning the niche of dark diversity. *Trends in Ecology & Evolution*, 26, 265–266.
- Pärtel, M., Szava-Kovats, R. & Zobel, M. (2013) community completeness: linking local and dark diversity within the species pool concept. *Folia Geobotanica*, 48, 307–317.
- Ricklefs, R.E. (1987) Community diversity – Relative roles of local and regional processes. *Science*, 235, 167–171.
- Ronk, A., Szava-Kovats, R. & Pärtel, M. (2015) Applying the dark diversity concept to plants at the European scale. *Ecography*, 38, 1015–1025.
- Tamme, R., Götzenberger, L., Zobel, M., Bullock, J.M., Hooftman, D.A.P., Kaasik, A. & Pärtel, M. (2014) Predicting species' maximum dispersal distances from simple plant traits. *Ecology*, 95, 505–513.
- Zobel, M. (1997) The relative role of species pools in determining plant species richness: An alternative explanation of species coexistence? *Trends in Ecology & Evolution*, 12, 266–269.
- Zobel, M., Maarel, E. & Dupré, C. (1998) Species pool: the concept, its determination and significance for community restoration. *Applied Vegetation Science*, 1, 55–66.

Submitted: 06 April 2016

First decision: 08 June 2016

Accepted: 27 June 2016

Edited by Marten Winter